

**UNIVERSIDAD  
NACIONAL DEL  
CHIMBORAZO**

**DATOS DE PANEL**

Patricia Hernández,  
PhD

MODELIZACIÓN  
ECONÓMICA

# RUTA A SEGUIR EN EL CURSO

01

ANÁLISIS EXPLORATORIO /  
CONFIRMATORIO DE DATOS

02 MODELOS  
ECONOMÉTRICOS

## MODELOS MICRO

- Regresión lineal
- Variable dependiente cualitativa (probit / logit)
- Variable dependiente categórica (ordinal / multinomial)
- **Datos de panel**

# **MODELOS DE DATOS DE PANEL**

# LOS DATOS PARA EL PANEL

CORTE TRANSVERSAL + SERIES DE TIEMPO = DATOS DE PANEL



MCO  
Probit  
Logit



Vs



Vs



Vs



## CORTE TRANSVERSAL

ARIMA  
VAR

VEC / Cointegración



## SERIE DE TIEMPO

1970

1980

1990

2010

2020



# LOS DATOS PARA EL PANEL

Y {  
i componente individual (personas, países...)  
t componente temporal (anual, semestral....)



Yit

# LOS DATOS PARA EL PANEL

ESTUDIOS DE SEGUIMIENTO

VARIABLE: ALTURA



0,60 m

1 m

1,30 m

1,60 m

1,68 m

# LOS DATOS PARA EL PANEL

ESTUDIOS DE SEGUIMIENTO

VARIABLE: ALTURA



0,70 m

1,10 m

1,35 m

1,70 m

1,75 m

# LOS DATOS PARA EL PANEL

<b>t</b>	<b>Id</b>	<b>Ingreso</b>	<b>Sexo</b>
To	A	100	M
T1	A	250	M
T2	A	500	M
To	B	150	F
T1	B	200	F
T2	B	300	F
To	C	500	M
T1	C	1000	M
T2	C	1200	M

- Formato largo de datos de panel
- Hay variables que no cambian en el tiempo (sexo)
- Se puede ordenar por tiempo o por individuo

# LOS DATOS PARA EL PANEL

p.e., un panel balanceado				p.e., un panel NO balanceado				
individuo	año	renta	edad	individuo	año	renta	edad	sexo
1	2000	1800	29	1	2000	800	19	2
1	2001	1950	30	1	2001	950	20	2
2	2000	800	20	2	2000	1900	29	1
2	2001	850	21	2	2001	1950	30	1
				2	2002	2100	31	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
500	2000	2200	54	1000	2000	2100	49	1
500	2001	2400	55	1000	2001	2200	50	1

# LOS DATOS PARA EL PANEL

Para **paneles balanceados**, describir el número de observaciones implica:

- Número de individuos distintos:  $N$
- Total de periodos cubiertos por el panel : $T$
- El número total de observaciones es simplemente  $NT$

Para **paneles NO balanceados**, además debemos considerar:

- Periodos concretos en que se observa cada individuo  $T_i$  (o su media)
- Número total de observaciones  $\sum T_i$

**¿SE PUEDE EMPLEAR MCO?**

**NO**

**¿POR QUÉ?**

**NO RECOGE LAS FUENTES DE VARIABILIDAD O  
HETEROGENEIDAD**

# FUENTES DE HETEROGENEIDAD

FUENTES DE VARIABILIDAD O HETEROGENEIDAD



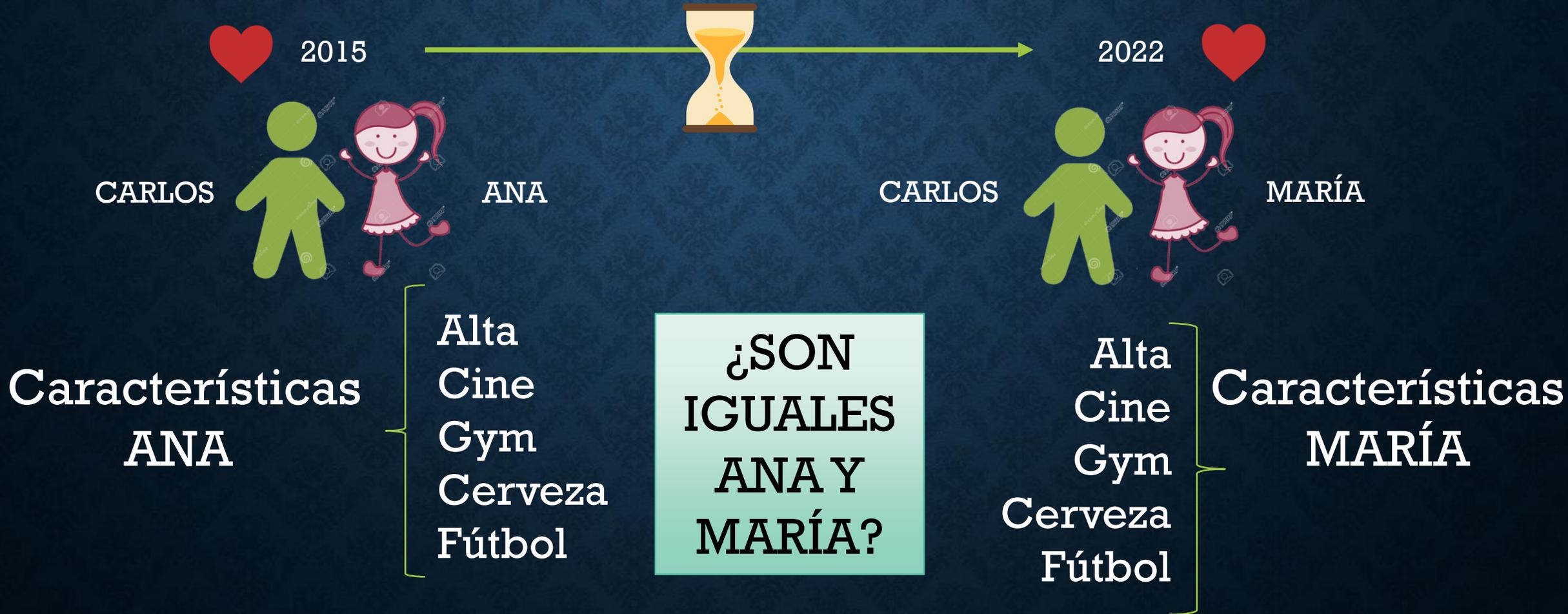
INDIVIDUOS - BETWEEN



TIEMPO - WITHIN

# HETEROGENEIDAD NO OBSERVADA

## FUENTES DE VARIABILIDAD O HETEROGENEIDAD



# HETEROGENEIDAD NO OBSERVADA

¿SON IGUALES ANA Y MARÍA?

NO

Hay características no observables (Heterogeneidad)

TÓXICA  
Y QUÉ?



ANA

NON TOXIC



MARÍA

$\alpha_i$

Heterogeneidad no observada

# MODELOS QUE RECOGEN HETEROGENEIDAD

$$y_{it} = \beta_1 x_{1it} + \dots + \beta_k x_{kit} + u_{it}$$

$$y_{it} = \beta_1 x_{1it} + \dots + \beta_k x_{kit} + \alpha_i + \varepsilon_{it}$$

Donde:

$x_{kit}$  son las variables explicativas (observadas)

$u_{it} = \alpha_i + \varepsilon_{it}$  término de error compuesto (inobservado)

$\alpha_i$  efectos individuales (heterogeneidad no observada permanente en el tiempo)

$\varepsilon_{it}$  error ruido blanco

# MODELOS QUE RECOGEN HETEROGENEIDAD

$$y_{it} = \beta_1 x_{1it} + \dots + \beta_k x_{kit} + \alpha_i + \varepsilon_{it}$$

NO PUEDE ESTIMARSE PORQUE:

$$\text{Correlación } (X_{it}, \alpha_i) \neq 0$$

LA ESTIMACIÓN ES SESGADA

# MODELOS QUE RECOGEN HETEROGENEIDAD



MODELO DE  
DATOS  
AGRUPADOS  
POOLED

NO CONSIDERA NI  
A LOS INDIVIDUOS  
NI AL TIEMPO

NO RECOGE  
HETEROGENEIDAD

# MODELOS QUE RECOGEN HETEROGENEIDAD



## MODELO DE EFECTOS FIJOS

Estimación de la diferencia de cada variable con respecto a su media

La heterogeneidad proviene de los individuos, asume *Correlación*  $(X_{it}, \alpha_i) \neq 0$

El efecto de heterogeneidad no observada desaparece porque se resta con respecto a su media y son iguales

**Problema: se pierde la información de la variable  $x$  que no cambia en el tiempo.**

# MODELOS QUE RECOGEN HETEROGENEIDAD

## EFFECTOS FIJOS

$$y_{it} - \bar{y}_i = \beta_1(x_{1t} - \bar{x}_1) + \dots + \beta_k(x_{kt} - \bar{x}_k) + (u_{it} - \bar{u}_i)$$

$$u_{it} - \bar{u}_i = \underbrace{\alpha_i - \bar{\alpha}_i}_{\alpha_i} + \varepsilon_{it} - \bar{\varepsilon}_{it}$$

$\alpha$  desaparece porque  $\alpha_{it}$  – la media de  $\alpha$  es cero (Se elimina el componente heterogéneo)

# MODELOS QUE RECOGEN HETEROGENEIDAD



## MODELO DE EFECTOS ALEATORIOS

Estimación de la diferencia de cada variable con respecto a un factor por su media

La heterogeneidad proviene del azar, asume *Correlación*  $(X_{it}, \alpha_i) = 0$

El efecto de heterogeneidad no depende de los individuos por lo tanto no está correlaciona con X y no hay problemas en la estimación

No se pierde la información de las variables que no cambian en el tiempo

# MODELOS QUE RECOGEN HETEROGENEIDAD

## EFFECTOS ALEATORIOS

$$y_{it} - \lambda \bar{y}_i = (1 - \lambda)a + \beta_1(x_{1t} - \lambda \bar{x}_1) + \dots + \beta_k(x_{kt} - \lambda \bar{x}_k) + v_{it}$$

$$v_{it} = (1 - \lambda)\alpha_i + (\varepsilon_{it} - \lambda \varepsilon_{it})$$

$\lambda = 0$  es un pooled, si  $\lambda = 1$  es efectos fijos

$\lambda$  Está relacionada con el cociente de la varianza del error ruido blanco y la varianza de la heterogeneidad. Si la varianza de la heterogeneidad es grande el valor se acercará a 1 y la estimación está más relacionada con efectos fijos

# EFECTOS FIJOS VS ALETORIOS

## TEST DE HAUSMAN

Ho:  $\beta$  de efectos aleatorios =  $\beta$  efectos fijos (si se acepta se emplean efectos aleatorios)

H1:  $\beta$  de efectos aleatorios  $\neq$   $\beta$  efectos fijos (si se acepta se emplean efectos fijos)

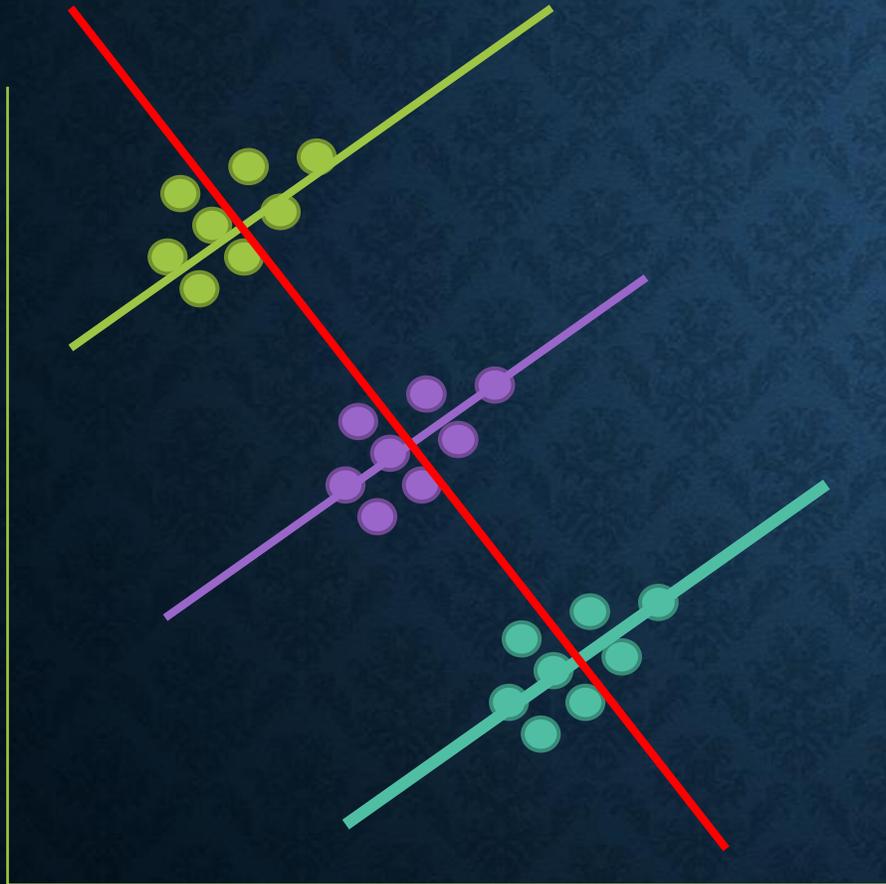
# EFECTOS FIJOS VS ALETORIOS

## TEST DE HAUSMAN

Ho:  $\beta$  de efectos aleatorios =  $\beta$  efectos fijos (si se acepta se emplean efectos aleatorios)

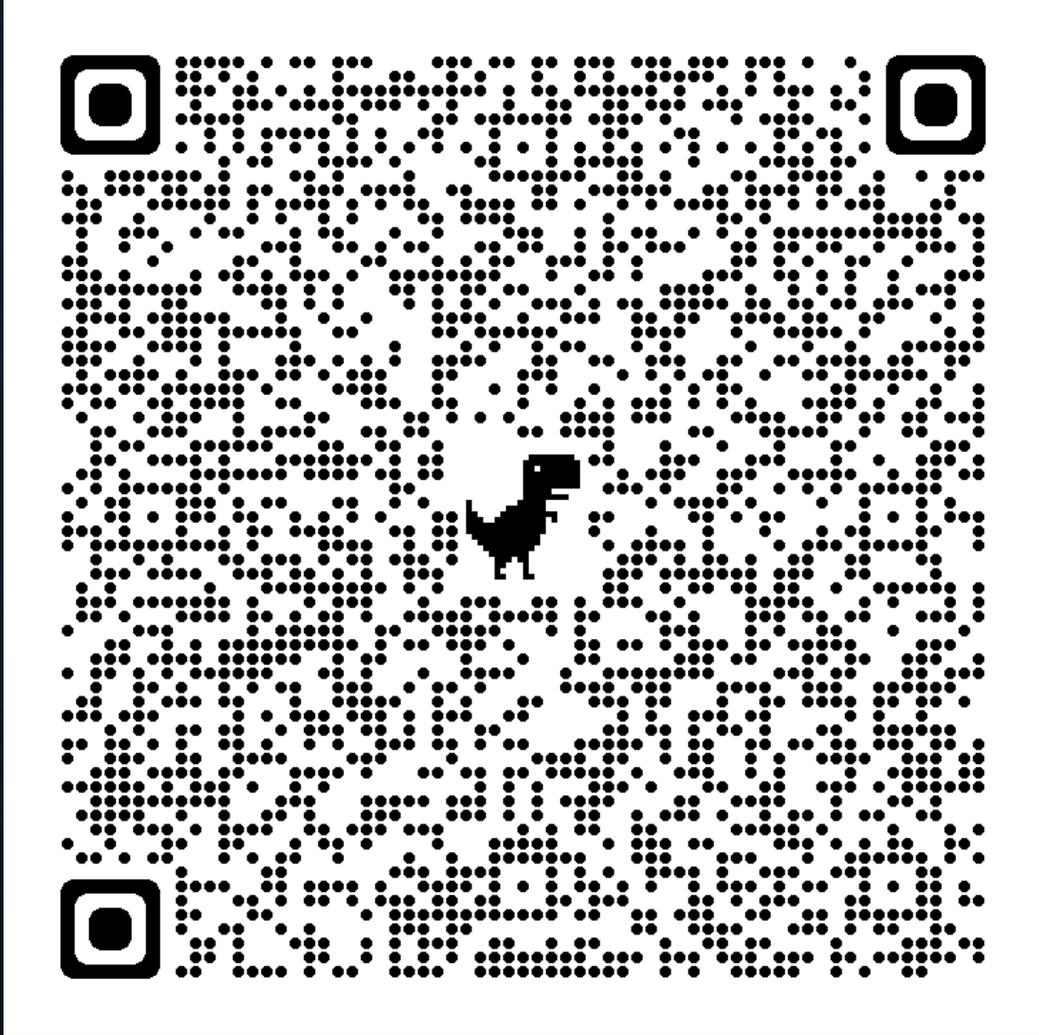
H1:  $\beta$  de efectos aleatorios  $\neq$   $\beta$  efectos fijos (si se acepta se emplean efectos fijos)

# EFFECTOS FIJOS VS ALEATORIOS



Cuando los  $\beta$  de efectos aleatorios (línea roja)  $\neq$   $\beta$  efectos fijos (Líneas de colores)

Se prefiere efectos fijos porque recoge las diferencias a pesar de perder la información de variables fijas



**APLICACIONES  
EN STATA  
(EJERCICIO  
5\_1)**

# ¿CÓMO EVALUAMOS AUTOCORRELACIÓN Y HETEROCEDASTICIDAD?



**AUTOCORRELACIÓN**

El error se relaciona con sus rezagos, pues recoge el componente del pasado de Y que no se incorpora en el modelo  
**COMPONENTE TEMPORAL**

**HETEROCEDASTICIDAD**

La varianza de los errores no es constante por problemas de especificación del modelo o por valores atípicos

# ¿CÓMO SE DETECTA Y CORRIGE?

## AUTOCORRELACIÓN

### SE DETECTA

Test de autocorrelación - Wooldridge  
Stata: Instalar xtserial  
Ho: no existe autocorrelación

## HETEROCEDASTICIDAD

### SE DETECTA

(Solo para EF porque RE se estima por GLS)

Test heterocedasticidad - Test modificado de White  
Stata: Instalar xttest3  
Ho: no hay heterocedasticidad

# ¿CÓMO SE DETECTA Y CORRIGE?

**AUTOCORRELACIÓN**

xtpcse var1 var2 var3, c(ar1) - Robusto  
xtgls var1 var2 var3, c(ar1) -GLS

**HETEROCEDASTICIDAD**

xtpcse var1 var2 var3 - robusto  
xtgls var1 var2 var3, p(h) -GLS

**HETERO + AUTO**

xtpcse var1 var2 var3, het c(ar1)  
xtgls var1 var2 var3, p(h) c(ar1)

**APLICACIONES  
EN STATA  
(EJERCICIO  
5\_1)**

# MODELOS DE DATOS DE PANEL DINÁMICO

# PANEL CON VARIABLES INSTRUMENTALES

$$y_{it} = W_{it}\gamma + X_{it}\beta + \alpha_i + \varepsilon_{it}$$

Donde:

$W_{it}$  son las variables explicativas ENDÓGENAS

$X_{it}$  son las variables explicativas EXÓGENAS

$\alpha_i$  efectos individuales (heterogeneidad no observada permanente en el tiempo)

$\varepsilon_{it}$  error ruido blanco

PROBLEMA

LAS VARIABLES EXPLICATIVAS ENDÓGENAS

# PANEL CON VARIABLES INSTRUMENTALES

$$y_{it} = W_{it}\gamma + X_{it}\beta + \alpha_i + \varepsilon_{it}$$

## CORRELACIONES EXISTENTES

$$\begin{aligned} \text{Corr}(W_{it}, \alpha_i) &\neq 0 \\ \text{Corr}(X_{it}, \alpha_i) &\neq 0 \end{aligned}$$



No hay problema porque es la heterogeneidad que se trata de resolver con el panel

$$\text{Corr}(W_{it}, \varepsilon_{it}) \neq 0$$



PROBLEMA – Estimadores sesgados

$$\text{Corr}(X_{it}, \varepsilon_{it}) = 0$$



Debe ser así para garantizar que el  $E(\varepsilon_{it}|X) = 0$ , condición para ser insesgados

# PANEL CON VARIABLES INSTRUMENTALES

$$\text{Corr}(W_{it}, \varepsilon_{it}) \neq 0$$



SE RESUELVE A TRAVÉS DE VARIABLES INSTRUMENTALES



LA VARIABLE INSTRUMENTAL (IV) REQUIERE DOS CONDICIONES

$\text{Corr}(IV, \varepsilon_{it}) = 0$  Condición de exogeneidad

$\text{Corr}(IV, W_{it}) \neq 0$  Condición de relevancia

# PANEL DINÁMICO

$$y_{it} = \rho y_{it-1} + X_{it}\beta + \alpha_i + \varepsilon_{it}$$

Donde:

$y_{it-1}$  son las variables explicativas ENDÓGENAS rezagada

$X_{it}$  son las variables explicativas EXÓGENAS

$\alpha_i$  efectos individuales (heterogeneidad no observada permanente en el tiempo)

$\varepsilon_{it}$  error ruido blanco

# PANEL DINÁMICO

$$y_{it} = \rho y_{it-1} + X_{it}\beta + \alpha_i + \varepsilon_{it}$$

Correlacionados



Donde:

$y_{it-1}$  son las variables explicativas **ENDÓGENAS rezagada**

$X_{it}$  son las variables explicativas **EXÓGENAS**

$\alpha_i$  efectos individuales (heterogeneidad no observada permanente en el tiempo)

$\varepsilon_{it}$  error ruido blanco

# PANEL DINÁMICO

$$\text{Corr}(y_{it-1}, \varepsilon_{it}) \neq 0$$



SE RESUELVE A TRAVÉS DE VARIABLES INSTRUMENTALES



DOS METODOLOGÍAS:

- ANDERSON – HSIAO
- ARELLANO BOND

# PANEL DINÁMICO

$$\text{Corr}(y_{it-1}, \varepsilon_{it}) \neq 0$$

ANDERSON - HSIAO



- Se estima en dos etapas (xtivreg)
- Los instrumentos son las diferencias de los rezagos de y
- Por lo tanto los resultados son en diferencias

ARELLANO - BOND



- Se estima en mínimos cuadrados generalizados (GLS) o en dos etapas (ivreg)
- Los instrumentos son las diferencias de los rezagos de la variable y, x
- Los resultados son en niveles
- Es necesario validar la correcta especificación (estat sargan, Ho correcta especificación) y la autocorrelación (estat abond, Ho no auto)

# PANEL DINÁMICO

$$\text{Corr}(y_{it-1}, \varepsilon_{it}) \neq 0$$

ANDERSON - HSIAO



```
xтивreg var1 var2 var3 (1.var=l2.var1), fd
```

ARELLANO - BOND



```
Xtabond var1 var2 var3, lags (1) GLS
```

```
Xtabond var1 var2 var3, lags (1), twostep (dos etapas)
```

**PANEL DINÁMICO**  
**MUCHOS DATOS**  

---

**EN EL TIEMPO**  

---

# PANEL DINÁMICO MUCHOS DATOS TIEMPO

## PASOS

### ESTACIONARIEDAD DE LAS VARIABLES

Test de Levin – Lin – Chu

Ho: no estacionario (raíz unitaria)

Stata: xtunitroot llc var1

Si se acepta Ho entonces se diferencia, se vuelve a correr test

Se determina el orden de integración var1 I(d)

Si las variables son I(0):

- Estimar VAR (en niveles)
- VEC / Cointegración

Si las variables son I(d):

- Validar si cointegran (igual orden de integración y residuos ruido blanco). Si es así estimar: VEC / Cointegración
- Si no cointegran estimar VAR en diferencias (CUIDAR LA PÉRDIDA DE DATOS)

# PANEL DINÁMICO

## MUCHOS DATOS TIEMPO

### PASOS

Si las variables son  $I(0)$ :  
Estimar VAR (en niveles)  
Si no son  $I(0)$  debe  
estimarse en diferencias  
(d.var)



Instalar pVAR

1. Número de rezagos óptimos: `pvarsoc var1 var2`
2. Estimación `pvar var1 var2, lags (#)`
3. Estabilidad del modelo `pvarstable`
4. Test de causalidad de granger `pvargranger`
5. Función impulso respuesta (`pvarirf`)
6. Descomposición de varianza (`pvarfevd`)

# PANEL DINÁMICO

## MUCHOS DATOS TIEMPO

### PASOS

Cointegración



Instalar `xtcointtest` y `xtcointreg`

Estimar test de cointegración, opciones (Ho: no cointegran):

`Xtcointest kao var1 var2 var3`

`Xtcointestst pedroni var1 var2 var3`

Si se rechaza Ho:

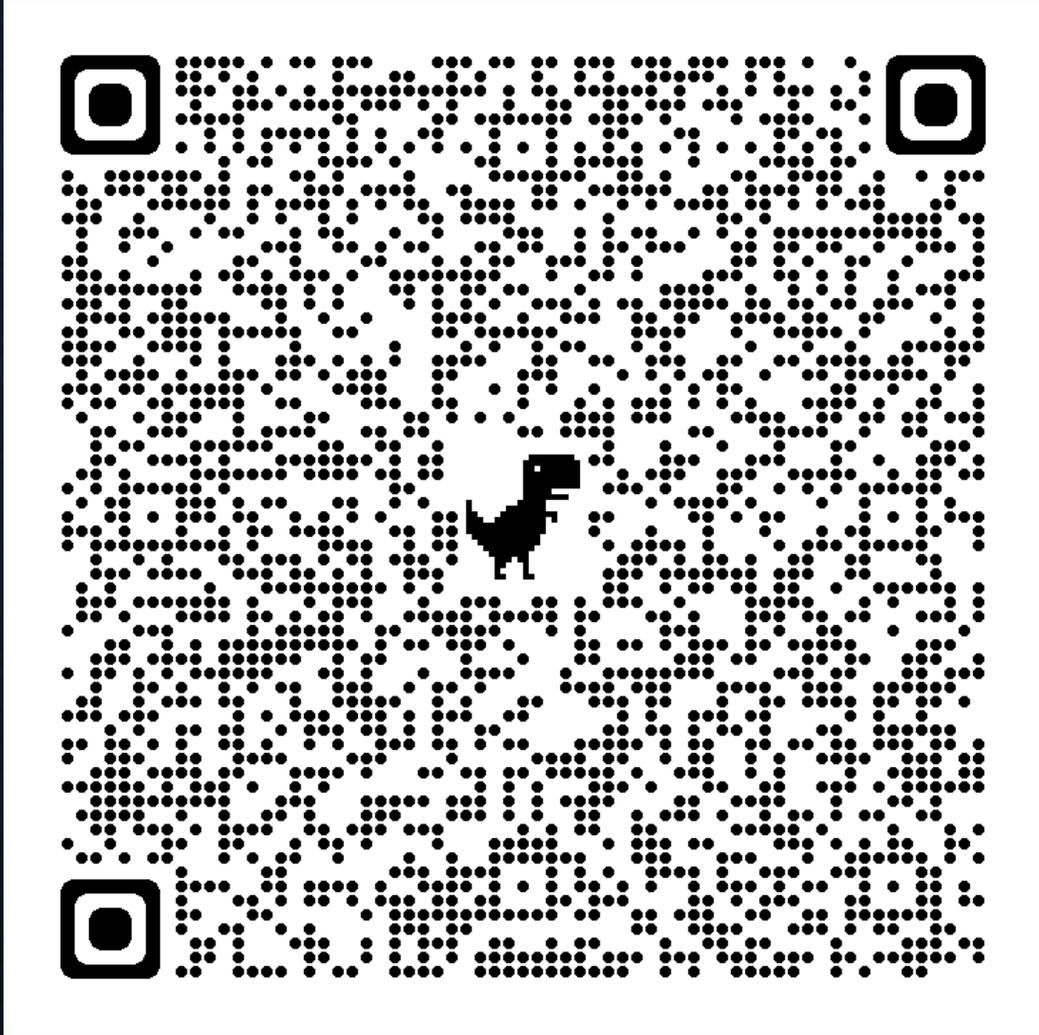
Estimar el modelo de cointegración

`Xtcointreg var1 var2 var3`

Si se acepta Ho:

Estimar el VAR en diferencias

(Pérdida de información)



**APLICACIONES  
EN STATA  
(EJERCICIO  
5\_2)**